# From Metadata to Meaning: GPT-4 Reveals Bias Trends in YouTube

Nitin Agarwal
COSMOS Research Center, University of Arkansas, Little Rock, USA
International Computer Science Institute, University of California, Berkeley, USA
e-mail: nxagarwal@ualr.edu

***Abstract*—YouTube's recommendation system significantly shapes user experiences but has raised concerns over potential bias and the formation of filter bubbles. Traditional studies have primarily relied on metadata, such as video titles, which often fail to capture the full context or nuance of video content. This study harnesses recent advancements in Artificial Intelligence (AI)—specifically the capabilities of Generative Pre-trained Transformer 4 (GPT-4)—to conduct a deep comparative analysis of sentiment, emotion, and toxicity across multiple layers of YouTube video content. By leveraging AI to extract and interpret narrative elements beyond superficial metadata, the research uncovers key patterns: a shift from neutral to positive sentiment and emotion (especially joy) with increased content depth, a consistent decrease in anger, and divergent toxicity trends—rising in titles but decreasing in deeper narrative analysis. These findings underscore AI's transformative role in enhancing content understanding and addressing long-standing challenges in recommendation system bias.**

***Keywords-YouTube recommendation system; artificial intelligence (AI); GPT-4; sentiment analysis; emotion analysis; toxicity analysis; bias; narrative analysis; recommender systems; social media algorithms; human-centered AI component; Open AI Whisper model.***

## I. Introduction

YouTube reports that 70% of user watch time is spent on recommended content. Powered by a multi-billion-dollar recommendation system, the platform drives average mobile viewing sessions beyond 40 minutes. This system forms a feedback loop: after a user selects a recommended video, new suggestions are generated, continuing the cycle. With over 2.7 billion users globally [1], and one-quarter of Americans using it as a primary information source [2], YouTube holds significant influence. This raises a crucial question: how much does the platform shape users' narratives?

Initially a video-sharing site, YouTube has evolved to maximize engagement, leveraging AI—particularly Google Brain—to personalize content delivery. This shift introduced algorithmic biases, such as selection bias [3], position bias [4][5], and popularity bias [6][7]. These biases contribute to "filter bubbles" and "echo chambers," where users are exposed to homogenous content, reinforcing existing views and limiting exposure to diverse perspectives.

Previous studies on YouTube's recommendation system have mostly relied on metadata like titles and descriptions. While useful, this approach often fails to capture the full context or narrative of the videos. Titles, crafted for brevity or clickbait, may not reflect the depth or tone of the actual content. This gap complicates content analysis and risks misrepresenting the emotional or toxic elements embedded within videos.

Our study addresses this gap by analyzing deeper content layers, focusing on sentiment, emotion, and toxicity. We use Generative Pre-trained Transformer 4 (GPT-4) to generate abstractive narrative summaries from video content, moving beyond surface-level metadata. This allows us to explore complex topics, such as the South China Sea dispute, through the lens of embedded narratives, diasporic perspectives, foreign policy, and global economic dynamics.

The goal is to understand how content evolves in depth and how it may influence recommendation patterns. We investigate whether deeper content carries different emotional or toxic tones compared to titles and descriptions. This approach helps reveal underlying shifts in content nature and user experience as influenced by AI-driven recommendations.

The remainder of the paper is structured as follows: Section II reviews related literature, Section III outlines our methodology, and the subsequent sections present the results and conclusions.

## II. Literature Review

This section reviews relevant literature on morality assessment, emotion detection, and bias in recommendation systems. Substantial research has addressed recommendation bias, particularly in areas, such as radicalization and the spread of misinformation and disinformation [8]. Studies have examined the emergence of homophilic communities within recommended video content and the factors driving their formation. Some have also identified coordinated behavior among YouTube commenters, potentially shaping user engagement and content visibility [9][10][11]. These findings reveal patterns of homogeneity, networked communities, and systemic bias within recommendation algorithms.

A key method for studying content evolution is "drift" analysis. O'Hare et al. [12] used a sentiment-tagged corpus to detect topic drift in texts, while Liu et al. [13] applied Latent Dirichlet Allocation (LDA) to monitor topic drift in micro-blogs. Venigalla et al. [14] examined emotional trends in India during COVID-19 using real-time Twitter data, presenting mood shifts through line graphs and radar maps

ISBN: 978-1-68558-330-9

over a defined period. In this study, we apply drift analysis to assess shifts in emotion and morality, aiming to uncover latent biases in YouTube's recommendation algorithm. This dual analysis provides a comprehensive view of how emotional and moral tones evolve within recommended content.

Prior research has also focused on how biases in algorithmic suggestions foster ideological clustering and the spread of uniform viewpoints [15]. These studies highlight the formation of like-minded groups and user-driven amplification of content through coordinated interactions in comment sections [16]. Understanding these behaviors is critical to identifying how bias reinforces content homogeneity and the influence of algorithmic curation.

To examine narratives within content, researchers have drawn on the field of Computational Narratology, which explores narrative structures from an algorithmic and information-processing perspective. This involves steps, such as preprocessing, parsing, identifying and linking narrative components, representing narratives, and evaluating them [17]. The rise of large pre-trained language models, like GPT-3, has revolutionized narrative extraction, enabling models to identify key features and execute varied tasks with minimal training—often requiring only a well-crafted prompt. Advancements like trainable continuous prompt embeddings have improved model performance, enhancing GPT and Bidirectional Encoder Representations from Transformers (BERT) accuracy by up to 80% [18].

Recent work has also advanced the understanding of figurative language across both discriminative and generative tasks, narrowing the gap between model output and human interpretation [19]. These developments are central to our study, which uses GPT-4 to analyze YouTube narratives beyond surface-level metadata, enabling deeper insights into the emotional and moral dimensions embedded in recommended content.

## III. METHODOLOGY

Next, we describe the research methodology, including data collection, transcript generation, narrative extraction, sentiment, emotion, and toxicity analysis.

### A. *Data Collection*

YouTube's recommendation algorithm is heavily influenced by a user's viewing history, resulting in highly personalized suggestions. To minimize personalization bias and maintain experimental control, we applied the following precautions during data collection:

1. All watch sessions were conducted without logging into a YouTube account.

2. A fresh browser instance was launched for each new recommendation depth.

3. Cookies were cleared between depths to avoid cross-session influence and ensure unbiased retrieval.

We collected data from YouTube's "watch-next" panel using this setup. To define our video corpus, we collaborated with subject matter experts in structured workshops to develop a targeted list of keywords related to the South China Sea Dispute. Table 1 provides these keywords. These keywords were then used to generate search queries, producing an initial set of seed videos.

TABLE I. SOUTH CHINA SEA DISPUTE KEYWORDS.

| |
|---|
| South China Sea, SCS dispute, SCS conflict, Nine-dash line, Maritime sovereignty, Territorial waters, EEZ, Exclusive Economic Zone, Freedom of navigation, UNCLOS, United Nations Convention on the Law of the Sea, China South China Sea, Philippines South China Sea, Vietnam South China Sea, Malaysia South China Sea, Indonesia Natuna Sea, Taiwan South China Sea, US Navy South China Sea, PLA Navy, People's Liberation Army Navy, South China Sea military drills, Naval exercises SCS, Militarization of islands, Artificial islands South China Sea, Spratly Islands conflict, Paracel Islands tension, Freedom of navigation operations, FONOP, Aircraft carrier SCS, Strategic waterway Asia-Pacific, Hague tribunal South China Sea, PCA ruling 2016, South China Sea arbitration, Maritime law dispute, Sovereignty claims Asia, ASEAN South China Sea, South China Sea diplomacy, Regional security Indo-Pacific, #SouthChinaSea, #SCSdispute, #MaritimeTensions, #FreedomOfNavigation, #StopAggression, #DefendSovereignty, #AsiaSecurity, #GeopoliticsAsia |

From these seeds, we extracted recommendations across three recursive depths, with each video in one tier serving as the basis for collecting recommendations in the next. This iterative process produced a dataset of 9,372 videos. The initial tier included the top 75 most viewed seed videos, and each successive depth expanded by a factor of five to capture a broader and more representative sample with varying video quality.

### B. *Transcript Generation*

#### 1) *Collecting Transcripts from YouTube*

Efficient caption collection from YouTube requires extracting available manual or automatic transcripts, excluding videos with mixed-language dialogue. However, the YouTube Data API v3 [20], while rich in video metadata, restricts caption access due to copyright and privacy concerns. Even with policy changes, issues like rate limits and API key requirements would remain.

To address this, the study employed the YouTube Transcript API by Jonas Depoix [21], which bypasses official restrictions by simulating YouTube's client-side HTTP requests, accessing captions without authentication. The retrieval strategy prioritized human-generated English captions, followed by auto-generated English, then non-English captions—favoring human-created and English-language transcripts for greater accuracy.

To improve processing efficiency, the Python ThreadPoolExecutor [22] from the concurrent.futures module was used. This allowed concurrent caption retrieval across multiple videos, significantly reducing time delays from network calls.

Despite this streamlined approach, limitations persisted. Some videos lacked captions due to creator choices, legal constraints, or difficulties in transcribing multilingual

ISBN: 978-1-68558-330-9

content. As a result, the study underscores the need for alternative transcription methods, such as audio-based transcription for videos with no available captions, ensuring comprehensive dataset coverage.

*2) Generating Transcripts Unavailable from YouTube*

To transcribe YouTube videos without captions, we utilized the OpenAI Whisper model [23], which is trained on 680,000 hours of diverse, multilingual data. Whisper uses an encoder-decoder Transformer architecture [24] to convert audio into text, excelling in recognizing varied accents and background noise. Although it doesn't always outperform task-specific models, Whisper's broad training makes it ideal for general transcription.

Due to Whisper's processing latency, we adopted faster-whisper [25], a high-performance CTranslate2-based reimplementation. To boost efficiency, we customized several parameters. Disabling word-level timestamps reduced unnecessary computation, and enabling the vad filter removed non-speech audio, shortening transcription time. We also turned off condition on previous text to treat audio chunks independently, enabling parallelization with minimal quality loss.

Key decoding parameters were adjusted: temperature was set to 0 for deterministic outputs, while beam size was reduced from 5 to 3 to explore fewer yet high-quality paths, balancing speed and accuracy. Additionally, we lowered the patience parameter from 1.0 to 0.9, slightly shortening the beam search duration.

To further enhance throughput, we used Python's ProcessPoolExecutor [26] for parallel processing across multiple GPUs. This approach allowed us to simultaneously download audio and generate transcriptions, maximizing computational efficiency.

In sum, our optimized pipeline—built on Whisper and faster-whisper, combined with parallel execution—achieved high transcription accuracy and speed, enabling scalable caption generation for large video datasets.

*3) Translating Transcripts to English*

To standardize transcriptions in non-English or mixed languages, we translated them into English to enhance usability. Although the YouTube Transcript API and Whisper offer translation features, we opted for tools specifically optimized for speed and quality. First, we used fasttext-langdetect by Facebook AI Research [27] to detect the language of each transcription. This model uses word embeddings and n-gram analysis for high-accuracy detection across numerous languages, allowing us to bypass translations for transcriptions already in English.

For non-English content, we employed the M2M100 model [28], a multilingual encoder-decoder developed by Facebook AI. M2M100 supports direct translation between 100 languages without requiring English as an intermediary. It is particularly effective at preserving meaning, even for less common language pairs, and can be run locally without the constraints of commercial APIs. Despite its slower CPU performance, M2M100's translation quality made it the preferred choice.

To address performance limitations, we leveraged Python's ProcessPoolExecutor and multi-GPU batch processing. This setup allowed us to divide workloads across GPUs efficiently, significantly accelerating the translation process while preserving high accuracy. These strategies enabled us to streamline transcription and translation workflows for multi-language video datasets effectively and reliably. A detailed performance analysis of this approach is presented in [29].

## C. Narrative Extraction

While much of the existing research focuses on analyzing YouTube's metadata to uncover user opinions, our approach advances this by extracting deeper narratives directly from YouTube video transcripts. Given that video lengths vary widely—from a few minutes to several hours—it poses a significant challenge to derive meaningful emotional content from such extensive transcripts. To address this, we utilized large language models like GPT-4 ("gpt-4-0125-preview"), which support up to 128,000 tokens, enabling us to process lengthy transcripts effectively.

To extract coherent and structured narratives from these transcripts, we designed specific prompts tailored for GPT-4. During generation, we configured the temperature parameter to 0, ensuring deterministic outputs—this means the model consistently generates the same result given the same input. Additionally, we set both the frequency penalty and presence penalty to 0, which helps avoid the inclusion of repetitive phrases in the generated narratives.

For brevity and focus, we limited the model's output using a max_tokens value of 25, enabling us to generate concise yet informative narrative summaries. This methodological setup allowed us to capture the embedded storytelling within the transcripts efficiently while maintaining consistency and clarity in the outputs generated by the language model.

## D. Sentiment Analysis

RoBERTa-based sentiment analysis leverages the RoBERTa architecture, an enhanced version of BERT designed for improved accuracy and efficiency. Trained on extensive datasets with labeled text, RoBERTa excels at identifying sentiment by capturing nuanced contextual cues within the input. This allows it to classify text into sentiment categories, such as positive, negative, or neutral, with high precision (94.2%). Its strong performance in understanding context has made it a popular choice for tasks like social media sentiment tracking and opinion analysis.

## E. Emotion Analysis

We analyzed the emotional content embedded in video-related text, including titles and extracted narratives, with a focus on seven key emotions: anger, disgust, fear, joy, neutral, sadness, and surprise. To identify emotional bias across various levels of video recommendations, we applied the concept of emotion drift. This drift was visualized using a line graph, where each point along the depth axis represented a different layer of the recommendation pathway. For improved accuracy in emotion classification, we employed a fine-tuned transfer learning model—

Emotion-English-DistilRoBERTa-base—tailored for Natural Language Processing (NLP) tasks.

### F. Toxicity Analysis

Detoxify, an open-source tool developed by Unitary AI [30], uses a Convolutional Neural Network (CNN) trained on word vector inputs to evaluate whether a given piece of text could be perceived as "toxic" within a conversational context. Upon receiving a text input, the Detoxify API generates a probability score between 0 and 1, with higher values indicating an increased likelihood of toxicity.

Detoxify provides toxicity scores across seven dimensions: overall toxicity (1), severe toxicity (2), obscenity (3), threats (4), insults (5), identity attacks (6), and sexually explicit content (7). The tool is particularly valuable due to its specialization in detecting harmful or inappropriate language online, making it a useful resource for analyzing user-generated content. Its accessibility as a Python library enhances its utility in research and application development aimed at moderating or understanding toxic discourse in digital environments.

## IV. Findings and Analysis

This section discusses our findings and their implications.

### A. Sentiment Analysis

Sentiment analysis of YouTube video content reveals a stark contrast between the emotional tone of titles and that of full narratives. Video titles generally exhibit a neutral sentiment, occasionally leaning positive as more descriptive language is used (see Figure 1). This neutral framing may be intentional—crafted to appeal broadly while concealing the more emotionally resonant elements of the content itself.

In contrast, the narratives embedded within video transcripts demonstrate a clear and consistent emotional progression, shifting from neutrality to distinctly positive sentiments over time (see Figure 2). This evolution suggests that the deeper emotional context and storytelling are largely reserved for the video's main content, rather than being reflected in the title. The result is a layered communication strategy in which the title functions as a broad hook, while the narrative delivers greater emotional nuance and engagement.

### B. Emotion Analysis

Figures 3 and 4 illustrate a clear emotional progression within YouTube video content, showing that as narrative depth increases, the expression of joy becomes more pronounced. This upward trend in positive emotion suggests a deliberate content strategy aimed at gradually eliciting stronger positive emotional responses from viewers. Simultaneously, there is a noticeable decline in negative emotions, such as anger, disgust, and sadness, indicating a possible intent to maintain viewer engagement through a more uplifting emotional arc.

When comparing emotional patterns between video titles and narratives, a distinct difference emerges. Titles tend to feature higher levels of negative emotions—especially disgust—at the outset. This may be a calculated use of emotional provocation or sensationalism to capture initial attention. However, as the content unfolds and narrative complexity increases, both titles and narratives begin to converge in emotional tone, trending more positively.

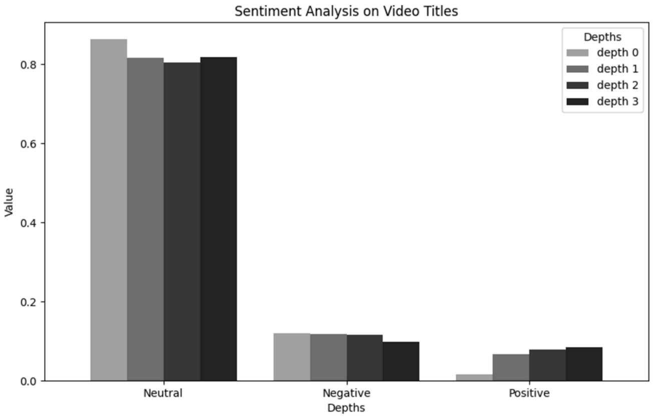

Figure 1. Sentiment trends for YouTube's video titles in different recommendation depths.

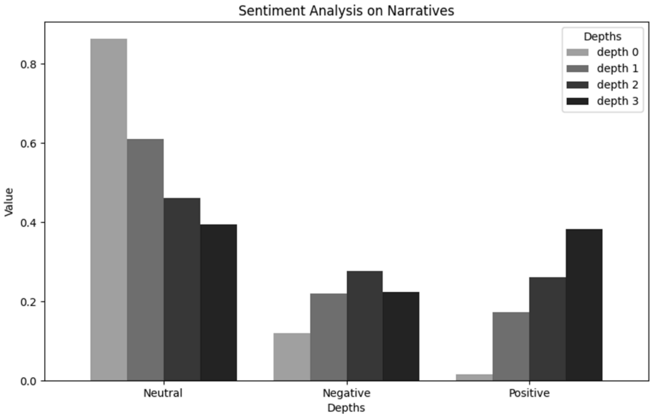

Figure 2. Sentiment trends for YouTube's video narratives in different recommendation depths.

This transition reflects a purposeful emotional modulation embedded in the content design. By initiating with emotionally charged titles and gradually shifting toward more positive sentiments in the narrative body, content creators appear to be leveraging emotion as a tool to sustain viewer interest while guiding the audience toward a more favorable emotional experience.

### C. Toxicity Analysis

As we explore successive layers of YouTube's recommendation system, the data reveals fluctuating yet overall increasing patterns in toxicity expressed in the video titles. Notably, an analysis of the corresponding video narratives shows a consistent decline in toxicity levels with each deeper level of recommendation. This downward trend in toxic content suggests that YouTube's algorithm may be effectively optimizing for more constructive or less harmful content over time. Figure 5 illustrates this progression, emphasizing how content becomes increasingly less toxic as users are guided further into the recommendation chain. This pattern points to a potential refinement in the platform's content curation strategy, aimed at fostering a healthier viewing environment.

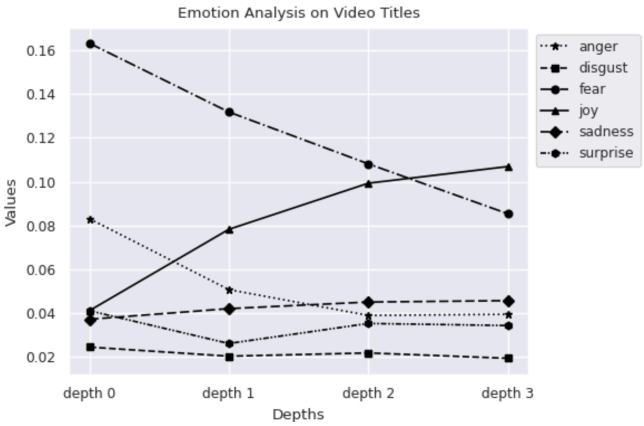

Figure 3. Emotion trends for YouTube's video titles in different recommendation depths.

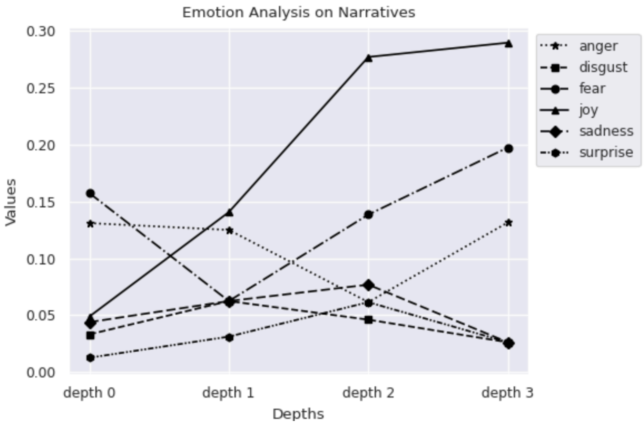

Figure 4. Emotion trends for YouTube's video narratives in different recommendation depths.

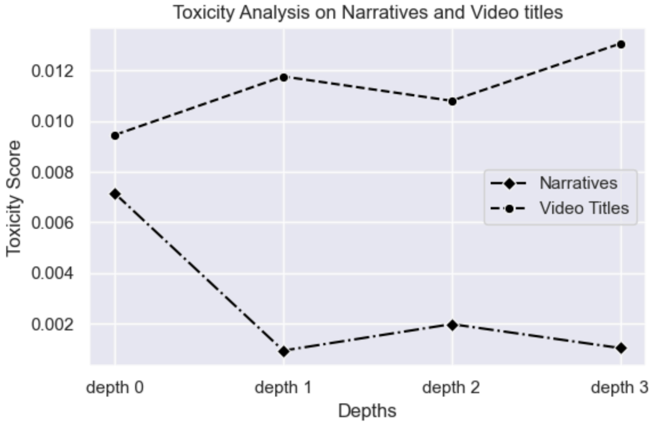

Figure 5. Toxicity trends for YouTube's video titles and narratives in different recommendation depths.

## V. Conclusion and Future Work

This study underscores the transformative role of AI in advancing the analysis of online video content. Traditional approaches that rely primarily on metadata, such as video titles, for sentiment and toxicity assessment are increasingly insufficient for capturing the nuanced emotional and behavioral dynamics present in digital media. Leveraging recent breakthroughs in AI, this research moves beyond surface-level indicators to extract, process, and interpret the rich narrative content found within YouTube videos.

By employing cutting-edge AI tools, such as large language models (e.g., GPT-4 for narrative extraction), fine-tuned transformer-based emotion classifiers (Emotion-English-DistilRoBERTa), and toxicity detection systems (Detoxify), this study taps into the full potential of modern NLP. These models enable the automated processing of large-scale video transcript data, allowing for the detection of subtle emotional patterns and toxic content that would be otherwise missed by metadata analysis alone.

The findings reveal that narratives contain more consistent and revealing emotional trajectories, particularly a notable increase in joy correlating with reduced toxicity, unlike the more volatile and superficial signals found in titles. These insights, powered by AI, advocate for a paradigm shift in content analysis methodologies. This research has several implications for media managers, strategic communications, content strategy, audience retention, ethical curation, competitive intelligence, and brand monitoring. Recommendations include moving beyond superficial engagement metrics, designing emotionally intelligent content strategies, ethically curating media to foster trust and long-term loyalty, and gaining a competitive edge by understanding how joy, coherence, and emotional authenticity drive success.

In essence, this research demonstrates how AI is not merely a tool but a driving force enabling deeper, more accurate, and scalable investigations into online media content. It highlights the need to embed AI-driven narrative analysis into future content moderation strategies and recommendation systems, setting a new standard for understanding the sentiment and toxicity landscape of platforms like YouTube.

## Acknowledgment

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920), U.S. Office of the Under Secretary of Defense for Research and Engineering (FA9550-22-1-0332), U.S. Army Research Office (W911NF-23-1-0011, W911NF-24-1-0078), U.S. Office of Naval Research (N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), U.S. Air Force Research Laboratory, U.S. Defense Advanced Research Projects Agency, the Australian Department of Defense Strategic Policy Grants Program, Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment, and the Donaghey Foundation at the University of Arkansas at Little Rock. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

## REFERENCES

[1] Statista Research Department, “YouTube users worldwide 2020-2029,” Statista, March 2025. [Online]. Available from: https://www.statista.com/forecasts/1144088/youtube-users-in-the-world [Last accessed: May 27, 2025].

[2] G. Stocking, P. Kessel, M. Barthel, K. Matsa, and M. Khuzam, “Many Americans Get News on YouTube, Where News Organizations and Independent Producers Thrive Side by Side,” Pew Research Center, September 2020. [Online]. Available from: https://www.pewresearch.org/journalism/2020/09/28/many-americans-get-news-on-youtube-where-news-organizations-and-independent-producers-thrive-side-by-side/ [Last accessed: May 27, 2025].

[3] Z. Ovaisi, R. Ahsan, Y. Zhang, K. Vasilaky, and E. Zheleva, “Correcting for selection bias in learning-to-rank systems," In Proceedings of The Web Conference 2020, pp. 1863-1873, 2020.

[4] A. Agarwal, I. Zaitsev, X. Wang, C. Li, M. Najork, and T. Joachims, “Estimating position bias without intrusive interventions,” In Proceedings of the twelfth ACM international conference on web search and data mining, pp. 474-482, 2019.

[5] X. Wang, N. Golbandi, M. Bendersky, D. Metzler, and M. Najork, “Position bias estimation for unbiased learning to rank in personal search,” In Proceedings of the eleventh ACM international conference on web search and data mining, pp. 610-618, 2018.

[6] R. Cañamares and P. Castells, “Should I follow the crowd? A probabilistic analysis of the effectiveness of popularity in recommender systems,” In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 415-424, 2018.

[7] H. Abdollahpouri, R. Burke, and B. Mobasher, “Controlling popularity bias in learning-to-rank recommendation,” In Proceedings of the eleventh ACM conference on recommender systems, pp. 42-46, 2017.

[8] M. Faddoul, G. Chaslot, and H. Farid, “A longitudinal analysis of YouTube's promotion of conspiracy videos,” arXiv preprint arXiv:2003.03318, 2020.

[9] S. Shajari and N. Agarwal, “Safeguarding YouTube Discussions: A Framework for Detecting Anomalous Commenter and Engagement Behaviors,” Journal of Social Network Analysis and Mining (SNAM), vol. 15, no. 54, pp. 1-24, Springer, 2025, DOI: 10.1007/s13278-025-01470-7.

[10] S. Shajari and N. Agarwal, “Developing a Network-Centric Approach for Anomalous Behavior Detection on YouTube,” Journal of Social Network Analysis and Mining (SNAM), vol. 15, no. 3, pp. 1-16, Springer, 2025, DOI: 10.1007/s13278-025-01417-y.

[11] S. Shajari, M. Alassad, and N. Agarwal, “Commenter Behavior Characterization on YouTube Channels,” In Proceedings of the Fifteenth International Conference on Information, Process, and Knowledge Management (eKNOW 2023), pp. 59-64, Venice, Italy, April 24 – 28, 2023.

[12] N. O'Hare, M. Davy, A. Bermingham, P. Ferguson, P. Sheridan, C. Gurrin, and A. Smeaton, “Topic-dependent sentiment analysis of financial blogs,” In Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion, pp. 9-16, 2009.

[13] Q. Liu, H. Huang, and C. Feng, “Micro-blog post topic drift detection based on LDA model,” In International Workshop on Behavior and Social Informatics and Computing, pp. 106-118, Cham: Springer International Publishing, 2013.

[14] A. Venigalla, S. Chimalakonda, and D. Vagavolu, “Mood of India during Covid-19-An interactive web portal based on emotion analysis of Twitter data,” In Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing, pp. 65-68, 2020.

[15] A. Chaney, B. Stewart, and B. Engelhardt, “How algorithmic confounding in recommendation systems increases homogeneity and decreases utility,” In Proceedings of the 12th ACM conference on recommender systems, pp. 224-232, 2018.

[16] K. Hosanagar, D. Fleder, D. Lee, and A. Buja, “Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation,” Management Science, vol. 60, no. 4, pp. 805-823, 2014.

[17] B. Santana et al., “A survey on narrative extraction from textual data,” Artificial Intelligence Review, vol. 56, no. 8, pp. 8393-8435, 2023.

[18] D. Stammbach, M. Antoniak, and E. Ash, “Heroes, villains, and victims, and GPT-3: Automated extraction of character roles without training data,” arXiv preprint arXiv:2205.07557, 2022.

[19] X. Liu et al., “GPT understands, too,” AI Open, vol. 5, pp. 208-215, 2024.

[20] Google Developers, “YouTube Data API,”. Google, n.d. [Online]. Available from: https://developers.google.com/youtube/v3 [Last accessed: May 27, 2025].

[21] J. Depoix, “youtube-transcript-api,” GitHub. [Online]. Available from: https://github.com/jdepoix/youtube-transcript-api [Last accessed: May 27, 2025].

[22] Python Software Foundation, “ThreadPoolExecutor,” in concurrent.futures — Launching parallel tasks, Python 3 documentation. [Online]. Available from: https://docs.python.org/3/library/concurrent.futures.html [Last accessed: May 27, 2025].

[23] A. Radford et al., “Robust Speech Recognition via Large-Scale Weak Supervision,” in Proceedings of the 40th International Conference on Machine Learning, vol. 202, pp. 28492-28518, July 2023.

[24] A. Vaswani et al., “Attention is all you need,” Advances in Neural Information Processing Systems, vol. 30, 2017.

[25] SYSTRAN, “faster-whisper,” GitHub. [Online]. Available from: https://github.com/SYSTRAN/faster-whisper [Last accessed: May 27, 2025].

[26] Python Software Foundation, “ProcessPoolExecutor,” in concurrent.futures — Launching parallel tasks, Python 3 documentation. [Online]. Available from: https://docs.python.org/3/library/concurrent.futures.html [Last accessed: May 27, 2025].

[27] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of Tricks for Efficient Text Classification,” arXiv preprint arXiv:1607.01759, 2016.

[28] A. Fan et al., “Beyond English-centric multilingual machine translation,” Journal of Machine Learning Research, vol. 22, no. 107, pp. 1-48, 2021.

[29] M. Cakmak and N. Agarwal, “High-speed transcript collection on multimedia platforms: Advancing social media research through parallel processing,” In 2024 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp. 857-860, IEEE, 2024.

[30] Unitary AI, “Unitary Virtual Agents: Effortless Automation with Human-level Precision,” [Online]. Available from: https://github.com/unitaryai/detoxify [Last accessed: June 14, 2025].

ISBN: 978-1-68558-330-9